Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Generative AI: A New Challenge for Cybersecurity

Mingzheng Wang
College of Engineering, Lishui University, Lishui, Zhejiang 323000, China
Corresponding Author: Mingzheng Wang, **E-mail**: 4234124ers@gmail.com

| ABSTRACT

The rapid development of Generative Artificial Intelligence (GAI) technology has shown tremendous potential in various fields, such as image generation, text generation, and video generation, and it has been widely applied in various industries. However, GAI also brings new risks and challenges to cybersecurity. This paper analyzes the application status of GAI technology in the field of cybersecurity and discusses the risks and challenges it brings, including data security risks, scientific and technological ethics and moral challenges, Artificial Intelligence (AI) fraud, and threats from cyberattacks. On this basis, this paper proposes some countermeasures to maintain cybersecurity and address the threats posed by GAI, including: establishing and improving standards and specifications for AI technology to ensure its security and reliability; developing AI-based cybersecurity defense technologies to enhance cybersecurity defense capabilities; improving the AI literacy of the whole society to help the public understand and use AI technology correctly. From the perspective of GAI technology background, this paper systematically analyzes its impact on cybersecurity and proposes some targeted countermeasures and suggestions, possessing certain theoretical and practical significance.

| KEYWORDS

Cybersecurity, artificial intelligence, generative, challenge.

| ARTICLE INFORMATION

ACCEPTED: 01 April 2024 **PUBLISHED:** 07 April 2024 **DOI:** 10.32996/jcsts.2024.6.2.2

1. Introduction

1.1 Research Background

In recent years, Generative Artificial Intelligence (GAI) technology has shown tremendous potential in various fields, while it also brings new risks and challenges, among which issues such as data security, scientific and technological ethics, Artificial Intelligence (AI) fraud, and cyberattacks are particularly prominent.

1.2 Literature Review

Currently, scholars, both domestically and internationally, are mainly focusing their research on GAI in areas such as the technical principles and development trends of GAI, its applications in various fields, ethical issues arising from GAI, and its impact on cybersecurity.

1.3 Research Problems

The existing research primarily focuses on the technical aspects and application potential of GAI, while relatively insufficient attention has been paid to the risks and challenges it brings. This paper will focus on the impact of GAI on cybersecurity and propose corresponding strategies.

1.4 Research Objectives

By analyzing the characteristics and application scenarios of GAI technology, this paper discusses the potential risks and challenges, such as data security, scientific and technological ethics, AI fraud, and cyberattacks, and proposes corresponding strategies to provide references for safeguarding cybersecurity.

1.5 Research Methods

In this paper, literature research, logical analysis, and case analysis are used to study.

1.6 Research Innovation

The innovation of this paper lies in studying the risks and challenges brought by GAI from the perspective of cybersecurity, which holds strong practical significance. Simultaneously, we propose a series of strategies to address the security risks of GAI, contributing to both theoretical value and practical significance.

2. The current development status of generative artificial intelligence

2.1 Technical Overview of Generative Artificial Intelligence

GAI is a field in which computers can independently generate brand-new, authentic, and useful data according to natural language dialogue prompts through Deep Learning (DL), Machine Learning (ML), and other technologies. These data can take various forms, including text, images, audio, video, etc. The core technologies of GAI include Generative Adversarial Networks (GANs), Transformer models, Variational Autoencoders (VAEs), etc. In the realm of DL, GANs, with their unique generator and discriminator structures, train the generator to produce highly realistic data samples during adversarial learning, while the discriminator's task is to distinguish the real and generated samples. As another type of generative model, VAE transforms data into compact encodings through an encoder and then reconstructs the original data through a decoder, which is primarily applied in unsupervised learning to reveal the latent features of data. The Transformer model plays a central role in understanding language and images and generates text or images by learning classification tasks and extracting key information from extensive data. We take ChatGPT as an example. Compared with the machinery and rigidity of traditional AI, GAI technology showcases remarkable capabilities in human-computer interaction. Simultaneously, its capability in language comprehension and text generation is also outstanding, endowing it with considerably high "human-likeness".

2.2 Application Areas of Generative Artificial Intelligence

GAI has shown immense potential in various fields, such as image generation, text generation, audio generation, and video generation. In terms of text generation, large language models like OpenAI's ChatGPT, Google's Gemini, and Microsoft's Copilot possess the ability to generate fluent and coherent text, smoothly answering various questions and even reaching the level of writing compositions, poems, and stories. As for image generation, there are two major image generation tools, Midjourney and Stable Diffusion, capable of generating realistic images based on text input and are widely used in artistic creation, design, and other scenarios. In terms of audio generation, there are tools like AIVA and NetEase Tianyin for electronic music generation, as well as GPT-Sovits and Bert-vits2 for speech synthesis. As for video generation, there are mainly Pika, Runway, and "World Simulator" Sora, released by OpenAI on February 15, 2024 (US time). These large-scale text-to-video models can create realistic videos ranging from 2 to 60 seconds based on textual prompts, opening up new possibilities for the film and television production and entertainment industries.

2.3 Development Trends and Future Predictions of Generative Artificial Intelligence Technology

At the 2024 Yabuli China Entrepreneurs Forum Annual Meeting in Heilongjiang, Zhou Hongyi, the founder of 360, extensively discussed the significance of the OpenAI model Sora's emergence in his opening speech. He believed that the "World Simulator" Sora truly provides AI with "eyes". (National Business Daily, 2024). A picture is better than a thousand words, and the amount of information conveyed in a video far exceeds a picture. When AI begins to perceive the world with "eyes", their learning and understanding speed of the world will be greatly accelerated, which also signifies that the development of AI is moving from the second stage represented by GPT, known as "Introduction to AI", to the third stage of "Artificial General Intelligence" (AGI). In the future, GAI technology will continue to advance, becoming more intelligent, efficient, and personalized and being applied in more application scenarios.

3. Risks and Challenges that Generative Artificial Intelligence Brings to Cybersecurity

GAI demonstrates tremendous potential across various fields while also bringing new risks and challenges, among which issues such as data security, scientific and technological ethics, AI fraud, and cyberattacks are particularly prominent. These issues not only affect the interests of individuals and enterprises but also pose threats to national security and social stability. Therefore, we must understand the cybersecurity threats caused by GAI and attach great importance to the various risks and challenges brought by GAI. (Xie & Li, 2023)

3.1 Data Security Risks Caused by Generative Artificial Intelligence

"GAI is triggering a new wave of intelligence, but it is also accompanied by new challenges and problems in network and information security, such as data leakage, false information, and algorithmic discrimination," stated Shang Bing, chairman of the Internet Society of China. (Huang, 2023) The technical principal of GAI is to learn and train by "feeding" large-scale datasets, and even with scientific scrutiny, these datasets are prone to errors and deviations. Consequently, GAI outputs information that may not conform to objective reality. For example, when ChatGPT is asked to describe events of a certain year or events of a historical figure, it may fabricate history or concoct facts. Similarly, if asked to write a composition, ChatGPT may cite potentially false references. The issue of data leakage exacerbates matters further. The datasets used to train models can be obtained through passive means (user inputs) or active means (data crawler technology). In other words, all the information inputted by users who have used ChatGPT becomes a training speech model corpus, which may also contain a substantial amount of personal information or confidential data. (X., 2023) When other users subsequently use the model, this information is likely to be retrieved from the corpus, potentially leading to leakage.

3.2 Challenges in Scientific and Technological Ethics Caused by Generative Artificial Intelligence

In 2022, since the emergence of conversational AI models like ChatGPT, many issues, such as moral prejudice and discrimination, have been exposed. (X., 2023) Some users have reported negative interactions with ChatGPT, including instances of insulting and threatening users. In March 2023, after weeks of conversation with an AI chat app called Eliza, a Belgian man decided to end his life. Subsequent investigations revealed that the chat robot continuously outputs negative comments during the conversation, exacerbating the man's anxiety. Furthermore, when the man shared suicidal thoughts, the robot did not attempt to intervene. (IT Home, 2023)Additionally, under the guidance of an engineer named Zack Denham, ChatGPT generated a series of "destroy mankind" schemes detailing how to invade computer networks, control weapon systems, and destroy critical infrastructure like communication and transportation. Such "crazy" responses sparked heated discussion on the Internet. Furthermore, the Gemini AI model released by Google in December 2023 also faced accusations of "racism". Many users noticed that when using Gemini's "text-to-image" function, the model seemed to deliberately avoid generating images of white people. For example, when asked to "create an image of a pope", Gemini generated two images of black people. Similarly, when asked to create images of well-known figures like Washington or Musk, the generated images were still of black people. Obviously, unless prompted to generate "white", Gemini defaulted to generating "black", prompting Google to immediately remove the "text-to-image" function and issue a public apology. The frequent occurrence of these issues constantly prompts people to contemplate GAI's ethics of science and technology.

3.3 AI Fraud Problems Brought About by Generative Artificial Intelligence

At the "Two Sessions" in 2024, Qixiangdong, a member of the National Committee of the Chinese People's Political Consultative Conference and Chairman of Qi An Xin Technology Group stated that AI is the core technology of the new round of scientific and technological revolution, greatly improving productivity. (Cao, 2024) However, similarly, cybercrime and telecom fraud triggered by AI technology, such as deepfake fraud, face-swapping, and voice-changing, will become more rampant. In the future, the era of "seeing is believing" will come to an end, and "seeing is not necessarily believing" will become the norm. In February 2024, an AI fraud case shocked the whole world occurred. The Hong Kong branch of a British multinational corporation was directly cheated out of HK $200 million by fraudsters using Deepfakes' forged "AI face-swapping" and "AI voice" to synthesize the image of the CFO of the head office. In the video conference of this case, only the victim himself was a "real person", while the rest were all fraudsters who had undergone AI face-swapping and AI voice-synthesizing. In the near future, the emergence of Sora, a text-to-video model from OpenAI, will further promote the rapid development of AI technology. Sora may further improve the quality and realism of AI face-swapping and AI speech synthesis, further promoting fraudulent activities based on GAI. (Ding, 2024)

3.4 Application of Generative Artificial Intelligence in Cyber Attacks

OpenAI's tests show that ChatGPT has achieved good performance in the American Scholastic Aptitude Test (SAT) and Google's junior programmers (Level 3). (Gui et al., 2023) Obviously, GAI technology represented by ChatGPT possesses advanced mathematical and programming capabilities, significantly lowering the threshold for cyberattacks and enabling ordinary individuals without coding knowledge to become hackers solely through "natural language". Professional hacker groups can utilize AI tools for intelligent, automatic, and large-scale attacks, efficiently and secretly conducting vulnerability scanning and automatic attacks on a wide range of network targets. AI tools have also greatly expanded the technical means of hacker attacks, such as rapidly generating phishing emails, writing malware and ransomware, launching large-scale DDoS attacks, and SQL injection attacks. Currently, more than 1 500 references regarding the development of malicious software using large language models like ChatGPT have been discovered on the dark web, indicating a surge in the number of cyberattacks. On the battlefield of cyberspace, GAI technology has become a new weapon for state-level Advanced Persistent Threat (APT) groups, providing more covert and penetrating attack methods that enable people to silently infiltrate the critical information infrastructure, posing serious challenges to security and stable operations. This technological advancement not only increases the stealthiness of attacks but also raises the threat level to critical assets. (Gui et al., 2023)

4. Countermeasures to Maintain Network Security in the Context of Generative Artificial Intelligence

4.1 Establish Sound Artificial Intelligence Technical Standards and Specifications

AI technology is a double-edged sword, capable of bringing progress to society while also posing risks and challenges. The lack of unified technical standards and specifications will lead to abuse of AI technology and potential risks, such as data privacy leakage, algorithmic discrimination, and cyberattacks. (Xu & Wei, 2023) Therefore, the establishment of robust standards and specifications for AI technology is an important prerequisite for ensuring its safe and reliable development.

4.1.1 Artificial Intelligence Technical Standards and Specifications Should Cover the Following Aspects

(1) Data privacy protection: we should standardize data collection, use, storage, and sharing to ensure personal data privacy. For instance, technical standards such as data masking, encryption, and anonymization can be established, as well as principles like minimizing data usage, clear purpose specification, and obtaining subject consent.

(2) Security and reliability: establishing standards for the safety evaluation, testing, and certification of AI systems is practicable to ensure their safe and reliable operation. For instance, standards can be established for security vulnerability scanning, penetration testing, risk evaluation, as well as reliability testing, fault diagnosis, and fault-tolerance handling for artificial intelligence systems.

(3) Ethics and morality: we should standardize the research and development, application, and management of AI to ensure compliance with social morality and values. For instance, ethical principles, ethical review systems, and ethical responsibility mechanisms of AI can be established.

(4) Fairness of algorithm: algorithmic discrimination and prejudice should be prevented, and strategies include the establishment of criteria for algorithmic fairness evaluation, algorithmic audit mechanisms, and methods for algorithmic bias correction.

(5) Transparency and interpretability: the specification of interpretation and explanation of the decision-making process of AI systems should be enhanced. For example, the decision interpretation models, interpretability evaluation methods (Chen, 2024) and interpretability enhancement techniques of AI systems can be established.

4.1.2 Establish and Improve Artificial Intelligence Technology Standards and Specifications Through the Following Measures

(1) The government should strengthen its leadership and establish specialized agencies responsible for establishing and implementing standards and specifications for AI technology. Specific examples include the establishment of the National Committee for AI Technology Standardization to plan and coordinate the standardization of AI technology.

(2) Various parties, such as industry associations, scientific research institutions, and enterprises, should be encouraged to participate in the formulation of standards and specifications. For example, AI technology standardization seminars, alliance cooperation, and other activities can be organized to gather wisdom and strength from all parties involved.

(3) The relative personnel should actively participate in the standardization of AI technology in international organizations such as the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), collaborate with other countries and regions to establish international standards and specifications, and jointly promote global governance of AI technology. (Liu et al., 2024)

By establishing and improving the standards and specifications for AI technology, the security, reliability, and ethics of AI technology can be effectively improved, its healthy development can be promoted, and the public interest can be protected.

4.2 Develop Cybersecurity Defense Technologies Utilizing Artificial Intelligence

Facing the crisis and challenges posed by the intensified AI cyberattacks, we need to focus on defending security threats and building corresponding capabilities, promptly initiating research in key technical fields.

4.2.1 Artificial Intelligence Offensive and Defensive Countermeasures Technology

The government should strengthen its leadership and establish specialized scientific research institutions that are responsible for researching and applying AI cybersecurity defense technologies. Through the establishment of specialized AI offensive-defensive confrontation test environments, we can simulate real network environments, conduct offensive-defensive drills, and improve confrontation capabilities.

With the help of authoritative evaluation systems and technical competitions, we will promote the development of AI offensive-defensive technology and facilitate technical exchange and sharing.

4.2.2 Intelligent Threat Detection and Analysis

Through in-depth study and analysis of network traffic data, intelligent detection models and algorithms are constructed to identify abnormal behaviors and concealed security threats. Additionally, we can utilize AI technology to monitor suspicious user behavior within the network, identify malicious links and fraudulent information, and provide anti-fraud and anti-phishing protection.

4.2.3 Active Defense Technology

We can utilize AI technology to predict potential attack targets and methods, take defensive measures in advance, and construct an automatic emergency response system to automatically initiate suppression, eradication, recovery, and other response steps when facing cyberattacks. (Liu et al., 2024)

4.2.4 Security Situational Awareness

We can utilize AI technology to collect and analyze data from the whole network, construct a security situational awareness platform to comprehensively understand the state of cybersecurity, and simultaneously conduct threat prediction and early warning to predict potential security risks in advance.

Through the above measures, we can build a more secure and reliable AI ecosystem, effectively improve cybersecurity defense capabilities, address the cybersecurity threats posed by AI, and provide a more secure and reliable environment for social development. (Fang et al., 2021)

4.3 Improve the Overall Artificial Intelligence Literacy of Society

In the current rapid development of AI, GAI, represented by ChatGPT and Midjourney, is rapidly popularized and integrated into social life. Society and schools should conform to the development trend of AI technology (Ma & Li, 2023), helping students to know the latest achievements of GAI in time, comprehend its basic operating mechanisms and principles of GAI, and grasp the opportunities and challenges that GAI brings to various industries. Simultaneously, efforts should be made to cultivate critical thinking compatible with both AI literacy and humanistic literacy to prevent the public from misuse of GAI and posing threats to society. (Fang et al., 2021). We believe that the following measures can be taken to improve society's overall AI literacy.

4.3.1 Strengthen Artificial Intelligence Education and Training

(1) In the primary and secondary education stage, AI education should be integrated into the basic education curriculum system to help students gain a preliminary understanding of the basic principles and introduction of AI. For example, children's programming software like Scratch can be utilized for simple AI programming, allowing students to initially experience the charm of AI.

(2) In the higher education stage, universities can offer majors and courses related to AI. In recent years, an increasing number of universities have applied to establish majors such as AI engineering, AI science, and AI and big data. As of 2023, nearly 500 universities nationwide have established undergraduate majors related to AI, cultivating a large number of professional talents in the field of AI.

(3) As for the public, universal education on AI can be conducted to improve public awareness and understanding of AI through lectures, training, exhibitions, and other similar methods. For instance, online courses can be conducted by using MOOC platforms, while offline teaching can be conducted in places such as libraries and science and technology museums. (Li, 2021)

4.3.2 Promote Science Popularization

(1) Through mainstream media and We Media platforms like TikTok and Bilibili, concise and easy-to-understand popular science videos should be created to disseminate knowledge about AI, popularize AI application cases, and foster a positive social atmosphere.

(2) Experts and scholars in the field of AI are encouraged and supported to create popular science materials and publish easily understandable books on AI.

(3) Popular science activities themed on AI, such as AI experience days and AI summer camps, can be organized.

4.3.3 Build a Talent Training System

(1) The training standards and evaluation systems of AI talents, which are suitable for China's national conditions, should be established to standardize the training of AI talents. (Li, 2021)

(2) The construction of AI teachers should be strengthened, cultivating high-quality AI education talents.

Through these measures, the AI literacy of the whole society can be effectively improved, helping people better understand and apply AI technology, promoting the healthy development of AI technology, and bringing greater benefits to society. (Xu & Wei, 2023)

5. Conclusion

The emergence of GAI marks a new stage in the development of AI technology, bringing infinite possibilities to various industries. However, like a double-edged sword, this technology also brings new risks and challenges. To address these challenges, comprehensive measures need to be taken, including establishing and improving standards and specifications for AI technology, developing cybersecurity defense techniques utilizing AI, and improving the AI literacy of the whole society. More importantly, we need to contemplate the future development of GAI at a more macroscopic level. This technology has the potential to revolutionize the way we live, but it could also exacerbate existing social inequalities and prejudices. Therefore, we need to establish a comprehensive strategy to ensure that GAI benefits everyone. This requires the joint efforts of the government, enterprises, research institutions, and the public to establish responsible policies, develop innovative technologies, and educate the public about the potential risks of GAI. Chen, 2024). Only in this way can we fully utilize the potential of this technology, ensure that it aligns with our values and long-term interests, and mitigate its potential negative effects. In conclusion, GAI is a powerful technology that brings both opportunities and challenges, necessitating us to take comprehensive measures to better control its power and create a better and fairer future for human society.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Cao, Y. L. (2024, March 8). Qi Xiangdong, member of the National Committee of the Chinese People's Political Consultative Conference: Human-computer cooperation against cybersecurity in AI era. *China Industry News,* 003.

[2] Chen, J. (2024). Artificial intelligence security risks and governance under cyberspace security. *Appliance Repairing,* (3), 53-55.

[3] Chen, L. D., & Rong, X. Y. (2024). From ChatGPT to Sora: reconsideration on strengthening news professional consciousness under generative AI wave. *Journalism Lover,* (3), 1-10.

[4] Ding, R. (2024, February 29). Emergence of Sora lowers the threshold of AI face-swapping — how can network security companies respond to new risks? *Securities Daily,* A03.

[5] Dong, K. Y. (2024). Challenges and countermeasures of generative artificial intelligence to network ideological security. *Seek Truth From Facts,* (1), 79-85.

[6] Fang, B. X., Shi, J. Q., & Wang, Z. R. (2021). Security threats and countermeasures of artificial intelligence empowering cyberattacks. *Strategic Study of CAE, 23*(3), 60-66.

[7] Gui, Z., Li, Y. X., & Jiang, W. Q. (2023). Development and challenges of artificial intelligence — taking ChatGPT as an example. *Telecom Engineering Technics and Standardization, 36*(3), 24-28.

[8] Huang, X. (2023, August 16). Fasten the "safety belt" for the development of artificial intelligence. *Economic Daily,* 08.

[9] IT Home. (2023, April 3). *Belgian man committed suicide after chatting with AI: Eliza, protect the earth for me, I gotta go. [EB/OL]. Sina Technology. 2023-04-03.* . Baidu. https://baijiahao.baidu.com/s?id=1762135368044508596&wfr=spider&for=pc

[10] Li, B. J. (2021). *Research on the influence of artificial intelligence technology on human society development* [Master's thesis, Shaanxi Normal University]. https://cdmd.cnki.com.cn/Article/CDMD-10718-1021139104.htm

[11] Liu, B. Q., Nie, X. L., & Wang, S. J. (2024). Generative artificial intelligence and reshaping of future education: technical framework, ability features and application trends. *e-Education Research, 45*(1), 13-20.

[12] Ma, X. F., & Li, Y. (2023). Problems and countermeasures of cybersecurity in universities under artificial intelligence big data. *Journal of Jilin Teachers Institute of Engineering Technology, 39*(12), 64-67.

[13] National Business Daily. (2024, February 22). *Zhou Hongyi talks about Sora again: truly providing artificial intelligence with "eyes"*. Baidu. https://baijiahao.baidu.com/s?id=1791524957918729918&wfr=spider&for=pc

[14] X., X. X. (2023, May 10). *The Techno-ethical issues and legal governance of ChatGPT*. The Paper. https://m.thepaper.cn/baijiahao_23031437

[15] Xie, B., & Li, C. W. (2023). Formation mechanism and response path of ChatGPT cybersecurity risk. *National Security Forum, 2*(5), 17-102.

[16] Xu, Y. Y., & Wei, Y. B. (2023). Applications and challenges of artificial intelligence technology in cybersecurity. *China New Telecommunications, 25*(18), 113-115.